

Positionspapier: Multimodale Interaktionsanalyse

REINHOLD SCHMITT

1. MULTIMODALITÄT

Eine multimodale Konzeption von Interaktion ist in verschiedener Hinsicht die notwendige »Korrektur« einer wissenschaftshistorischen, aufnahme- und analysetechnisch bedingten Reduktion. Dies gilt für die technisch erstellten Interaktionsdokumente, welche die empirische Grundlage unterschiedlicher Formen von Interaktionsanalysen darstellen. Das gilt gleichermaßen für die mit der materialen Spezifik der Interaktionsdokumente zusammenhängende Komplexität der Untersuchungsgegenstände und –fragen. Schließlich gilt diese Einschätzung für die mit beiden Komplexitäten zusammenhängende methodische und konzeptionelle Adäquatheit.

Aus multimodaler Sicht erscheint die Perspektive, die durch die weitgehende bzw. ausschließliche Konzentration auf Verbalität als zentralem Untersuchungsgegenstand der empirischen Analyse von Interaktion entstanden ist, als monomodal. Diese Perspektive reflektiert – was die tatsächliche Komplexität der in den Interaktionsdokumenten festgehaltenen Ursprungssituation betrifft – weder einen theoretisch haltbaren, noch empirisch evidenten autonomen Status von Verbalität als dominanter Ausdrucksressource. Die theoretische, methodische und konzeptionelle Priorisierung von Verbalität entstand vielmehr im Zuge der Etablierung der konversationsanalytischen und gesprächsanalytischen Perspektive auf Interaktion. Sie hat ihren Ursprung in der »gründerzeitlichen« Relevantanzsetzung einer Situation, in der Beteiligte und Analytiker einen vergleichbar begrenzten Zugang zur faktischen, audiovisuellen Komplexität der Gesamtsituation hatten: dem Telefongespräch. Die Priorisierung von Verbalität macht in ihrer technisch motivierten Limitierung der Komplexitätserfassung bei einem solchen Untersuchungsgegenstand durchaus Sinn.

In dieser konversationsanalytischen Gründungsphase, die durch die Dominanz von Telefonkommunikation als dem zentralen Untersuchungsgegenstand gekennzeichnet ist, wurde die mit der Materialspezifik verbundene Reduktion der Komplexitätsrepräsentanz explizit reflektiert. Als die Konversationsanalyse dann jedoch die Telefonzelle verlies und die Tür zur Welt der audiovisuellen Face-to-face-Kommunikation mit Wahrnehmung-Wahrnehmungs-Relevanzen öffnete, wurde das Spannungsverhältnis zwischen verbaler Fixierung und audiovisueller Komplexität nicht mehr hinreichend methodologisch reflektiert. Dieser »wissenschaftsgeschichtliche

Reduktionismus« wurde auch im Verlaufe des späteren Siegeszuges der Konversationsanalyse und der Erweiterung der Untersuchungsgegenstände durch immer komplexere – beispielsweise institutionell geprägte – Interaktionssituationen nicht konsequent reflektiert oder konzeptuelle und methodisch umgesetzt.

Das Bewusstsein hinsichtlich der reduktionistischen Implikationen einer monomodal-verbalen Perspektive hat sich mit der flächendeckenden Substitution der ursprünglich rein auditiven Analysegrundlagen durch audiovisuelle Interaktionsdokumente grundlegend verändert. Dabei spielt auch die Herausbildung einer konsequent multimodalen Perspektive (mehrheitlich von Wissenschaftler/innen mit konversationsanalytischem Hintergrund) eine wesentliche Rolle, sowie die damit einhergehende »zweite Rezeption« von Ansätzen, die bereits sehr früh die faktische multimodale Komplexität des interaktiven Alltags berücksichtigt haben. Nicht zuletzt trägt hierzu jedoch die empirische Evidenz bei, die zeigt, dass und wie sich die online-analytische Wahrnehmung aktueller Sprecher/innen, in denen sie ihre interaktive Umgebung verstehensdokumentarisch auf ihre eigene momentane Beteiligung beziehen, auf die Äußerungsproduktion auswirkt.

1.2 Interaktionsanalytische Zugänge

Im Moment wird die aktuelle Entwicklung der empirischen Untersuchung von Kommunikation durch ein Kontinuum koexistierender Zugänge bestimmt, die sich bezüglich der theoretisch-methodisch-konzeptionellen Bedeutung von Verbalität unterscheiden. Diese kann man – unter Berücksichtigung dabei unvermeidbarer prototypischer Verzeichnung – wie folgt sehen:

1.3 Klassische Konversationsanalyse

Untersuchungen auf der Grundlage dieser Konzeption arbeiten nach wie vor mit Audioaufzeichnungen und gestalten ihre Gegenstandskonstitution so, dass Phänomene und Strukturen im Mittelpunkt des Erkenntnisinteresses stehen, die auf der Grundlage einer monomodalen und letztlich immer noch autonom gesetzten Ausdrucksressource, der Verbalität, untersucht werden.

1.4 Modalitätsspezifisch erweiterte Konversationsanalyse

Arbeiten mit diesem Schwerpunkt integrieren singuläre visuelle Ausdrucksressourcen bei der multimodalen Erweiterung verbal definierter Konzepte. Sie arbeiten daher auf der Grundlage audio-visueller Interaktionsdokumente. Das Erkenntnisinteresse wird auch hier weitgehend noch von Strukturen verbaler Kommunikation bestimmt. Es geht dabei beispielsweise um Fragen, wie verbale Mechanismen der Interaktionskonstitution wie etwa die Turn-Taking-Organisation durch Blickverhalten, Gestikulation, Mimik etc. mitbestimmt werden.

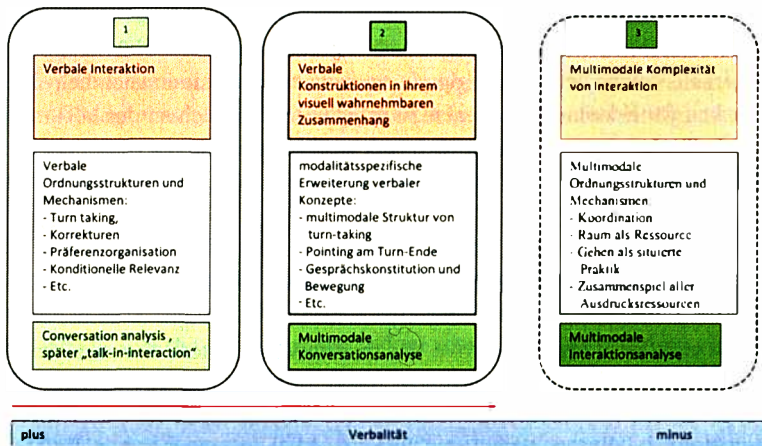
1.5 Multimodale Interaktionsanalyse

Dieser Ansatz benutzt ausschließlich Videoaufzeichnungen, hat aber im Unterschied zur multimodalen Konversationsanalyse per se kein primäres Interesse mehr an der Rekonstruktion ausschließlich verbaler Strukturen. Das Erkenntnisinteresse bezieht sich vielmehr auf die Rekonstruktion interaktiver Ordnungsstrukturen, die auf einer interaktionstheoretischen Grundlage arbeitet, für welche die Selbstbezeichnung in ihrer Kontrastivität gleichermaßen wissenschaftsgeschichtlich motiviert wie ein weißer Schimmel ist, denn Interaktion ist immer multimodal!

Zwischen diesen drei Formen konstitutionsanalytischer Rekonstruktion, die ihre gemeinsame methodologische Grundlage in der ethnomethodologischen Vorstellung einer »Vollzugsrealität« besitzen, besteht einerseits eine große Durchlässigkeit, andererseits gibt es auch deutliche Unterschiede. Letztere beziehen sich vor allem auf die theoretisch motivierte Art der Gegenstandskonstitution und auf die Rolle, welche die bei der Analyse verbaler Interaktion entwickelten Konzepte als Erkenntnisorientierung spielen.

Die relative Eigenständigkeit der Zugänge, ihre jeweiligen Schwerpunktsetzungen bei der Analyse von Interaktion, die Rolle, die Verbalität dabei spielt, sowie die Unterschiede in Methode, Theorie und der Bestimmung des Gegenstandsbereichs sind nochmals in folgender Grafik verdeutlicht.

Aktueller Stand: Empirische Interaktionsanalyse



Im Unterschied zu den beiden stärker verbal orientierten Zugängen ist für die multimodale Interaktionsanalyse charakteristisch, dass – ausgehend vom theoretischen Postulat der Egalität aller Ausdrucksressourcen – alle Formen interaktiver Praxis, ungeachtet der bei ihrer Konstitution eingesetzten Ausdrucksressourcen, gleichwertige Untersuchungsgegenstände sind. Das führt im Einzelfall (nicht grundsätzlich!) zur methodisch motivierten Fokussierung von Visualität als bislang weitgehend vernachlässigter Konstituente der Interaktion. Ihren methodischen Ausdruck findet dieses Verfahren beispielsweise in der »visuellen Erstanalyse« (Schmitt 2006, 2007a), bei der Verbalität zunächst motiviert ausgeblendet und erst im späteren Analysegang

integriert wird. Auch für die ko-konstruktive Gegenstandskonstitution öffnet sich dadurch der Blick auf Zusammenhänge, bei denen Verbalität nicht im Zentrum steht oder gar keine Rolle spielt. An die Stelle von Verbaltranskripten treten dann als empirische Basis der Analyse solcher Zusammenhänge neue Sekundärdokumente. Hierzu zählen beispielsweise Standbildreihen in Form von »Frame-Comics« (Schmitt/Dausendschön-Gay im Druck, Schmitt in Vorbereitung). Frame-Comics bilden den für die Analyse relevanten interaktiven Zusammenhang bewusst und motiviert in Form ausgewählter Standbilder ab, die entlang der faktischen Zeitlichkeit der Interaktion aus dem Primärdokument extrahiert werden.

Aus der Perspektive der multimodalen Interaktionsanalyse stellen sich auch die Fragen, in welcher Weise sich die Architektur von Räumen auf Interaktion auswirkt, welche Formen ermöglicht und welche eher verhindert werden und welches sozial geprägte, raumbezogenes Wissen sich in konkreten Formen der (vorinteraktiven) Raumnutzung manifestiert. Diese Fragen werden momentan unter den Begriffen „Interaktionsarchitektur“ und »Sozialtopografie« diskutiert (Hausendorf/Schmitt 2013).

Da sich die multimodale Interaktionsanalyse am Prinzip der methodischen Adäquatheit orientiert, steht die Entwicklung solcher Methoden und Konzepte im Vordergrund, die der interaktiven Komplexität durch die analytische Berücksichtigung aller Ausdrucksressourcen in ihrer sequenziellen und simultanen Vollzugscharakteristik und dem zwischen den einzelnen Ressourcen bestehenden Zusammenhang explizit Rechnung tragen. Damit ist zwangsläufig ein durch die Qualität des Untersuchungsgegenstandes motiviertes – produktiv zu verstehendes und zu nutzendes – »strukturelles Misstrauen« gegenüber monomodalen Konzepten verbunden. Im multimodalen Erkenntniszusammenhang entsteht grundsätzlich die Notwendigkeit, etablierte verbale Konzepte nicht fraglos zu nutzen, sondern sie hinsichtlich ihrer Angemessenheit für Erkenntnisinteressen zu reflektieren, die sich an der hör- und sichtbaren Komplexität von Interaktion und ihrer strukturimplikativen räumlichen Umgebung orientieren.

2. VISUELL WAHRNEHMBARE KO-KONSTRUKTIONEN

Interessiert man sich aus Perspektive der multimodalen Interaktionsanalyse für das breite Spektrum von Ko-Konstruktionen, dann bezieht diese Perspektive systematisch auch solche ko-konstruktiven Praktiken mit ein, für deren Konstitution Verbalität keine Rolle spielt. Für diese Perspektive sind vor allem zwei der von Dausendschön-Gay/Gülich/Krafft (i.d.Bd.) differenzierten Bereiche relevant, in denen die interaktionstheoretische Annahme des »*interactional achievement*« als erkenntnisleitender Rahmen für die konkrete Analyse ko-konstruktiver Praktiken analysefaktisch präzisiert wurde: Ko-Orientierung und Ko-Ordination (Dausendschön-Gay/Gülich/Krafft, i.d.Bd.: 34).

Dies betrifft insbesondere die für Ko-Orientierung konstitutive »Herstellung und Beibehaltung der gemeinsamen Aufmerksamkeitsausrichtung; die Organisation der Blickrichtung, deiktische Gestikulationen (»Zeigen«), sowie die für Koordination wesentlichen »Ereignisse der gemeinsamen Ausrichtung der Körperposituren, des Handlungsrythmus oder verschiedener Formen von »interaktivem Alignment« ... «.

3. GEMEINSAM GEHEN

Ein Prototyp solcher Ko-Konstruktionen bzw. ko-konstruktiver Praxis, die gänzlich ohne verbale Ko-Operation (dies ist der dritte in Dausendschön-Gay/Gülich/Krafft i.d.Bd. genannte Bereich) stellen die unterschiedlichsten, im Alltag omnipräsenten Formen des »gemeinsamen Gehens« dar.¹ Die ko-konstruktive Qualität des gemeinsamen Gehens (oder des Zusammen-Laufens) ist die wahrnehmbar gemachte, koordinierte, auf den jeweils anderen Beteiligten fein abgestimmte eigene Bewegung im Raum. Diese Bewegung unterscheidet sich erkennbar von der Art und Weise, in der Einzelpersonen – und auch nicht gemeinsam Gehende – die Distanz von A nach B zurücklegen. Der zentrale Unterschied besteht darin, dass beim gemeinsamen Gehen die sichtbare und symbolträchtige soziale Qualität der Bewegung einen eigenständigen Sinn bzw. Wert – eben den der Ko-Konstruktion erhält (siehe unten).

Ausgehend von einer Vorstellung, nach der es sich bei Ko-Konstruktionen um eine besondere Form der gemeinsamen Herstellung von Interaktion handelt, geraten all jene Formen interaktiver Praxis in den Blick, für die die Gemeinsamkeit der Orientierung, die Gleichsinnigkeit der Sinnproduktion, die kollektive Relevantsetzung thematischer oder pragmatischer Aspekte sowie eine wechselseitig akzeptierte koordinative Relevanz bestimmende Merkmale sind. Mechanismen der Interaktionskonstitution werden dazu genutzt, um gemeinsam eine spezifische Form sozialer Bedeutung zu schaffen, die durch ein kollektives Projekt getragen wird, welches die Bearbeiter auf spezifische Weise für die Dauer der Bearbeitung auch auf spezifische Weise sozial miteinander verbindet.

In einen solchen ko-konstruktiven Relevanzrahmen gehört – neben Ereignissen, für die Verbalität als konstitutiver Teil des sozialen Austauschs eine Rolle spielt – dann beispielsweise auch der gemeinsame Gang zweier Konfirmandinnen als Duett in der Eröffnungsphase eines Gottesdienstes (Schmitt 2012). Der einzige Unterschied zu verbalen Ko-Konstruktionen besteht darin, dass es sich dabei um »laufende Interaktionsbeiträge« handelt, durch die eine spezifische Form der Sinnkonstitution entsteht.

Um sich selbst und der Öffentlichkeit zu zeigen, dass sie als Duett im Kirchenraum unterwegs sind, realisieren die beiden Konfirmandinnen unterschiedliche Verfahren, die in ihrer empirischen Evidenz teils manifest, teils unauffällig sind. Grundsätzlich jedoch sind die Anforderungsstruktur und die zur Bearbeitung eingesetzten Verfahren strukturell mit verbalen Anforderungen und Bearbeitungsverfahren vergleichbar. Im Zentrum der von den Konfirmandinnen investierten Anstrengungen steht Koordination (Schmitt 2007b). Als Duett zu gehen bedeutet, die beiden individuellen Bewegungen kontinuierlich aufeinander abzustimmen und immer wieder mikrostrukturell zu überprüfen und zu justieren. Das, was das Video in Bezug auf den Gang der beiden Konfirmandinnen zeigt, ist nicht etwas fraglos Gegebenes oder Zufälliges. Es ist vielmehr das Ergebnis einer kontinuierlichen gemeinsamen Herstellung und die Folge der Bearbeitung interpersoneller koordinativer Relevanzen.

¹ Zur Bewegung im (öffentlichen) Raum – mit mehr oder weniger direktem Bezug auf den Aspekt der koordinierten gemeinsamen Hervorbringung – siehe beispielsweise die Untersuchungen von Ryave/Schenkein (1974), Collet/Marsh (1981) sowie Lee/Watson (1993).

Abbildung 1: Schritt- und Beinsynchronität bei Hin- und Rückweg zur Osterkerze



Der Gang der Konfirmandinnen zur Kerze ist ein Einzelfall und produziert zunächst fallspezifische Einsichten. Er ist jedoch gleichzeitig auch ein prototypischer Vertreter ›sprachfreier‹ Ko-Konstruktion (siehe auch Schmitt 2012b). In dieser Hinsicht ist er einerseits für die Entwicklung adäquater Analyseverfahren von wesentlicher Bedeutung. Andererseits erweitert er den Blick auf das Kaleidoskop ko-konstruktiver Vielfalt, die in keiner Weise an Verbalität gebunden ist. Die weitere Entwicklung multimodaler Analysen wird dazu beitragen, bislang aufgrund ihrer modalen Konstitutionsspezifik eher unbeachtete Interaktionsdokumente neben verbal konstituierte Ereignisse zu stellen und somit einen differenzierteren Einblick in das Universum ko-konstruktiver Strukturen zu eröffnen, die unseren sozialen Alltag bestimmen.

Einen noch wesentlich komplexeren Fall verbalitätsfreier Ko-Konstruktion stellt zum Beispiel der Umzug venezianischer Masken und Kostüme in einer provenzalischen Altstadt dar. Nur unter einer konsequent ko-konstruktiven Perspektive werden bei der Analyse sowohl der fein abgestimmte Mechanismus als auch die Mikroorganisation situativer Koordination deutlich, die sich im vielfältigen Zusammenspiel unterschiedlicher Beteiligungsweisen zeigen. Nur durch ko-konstruktive Aktivitäten kann ein soziales Ereignis im öffentlichen Raum entstehen. Ohne ko-konstruktive und koordinative Abstimmung hat man es lediglich mit merkwürdig gekleideten Personen zu tun, die durch die engen Straßen der Altstadt laufen. Unter einer ko-konstruktiven Sicht wird jedoch die beidseitige Verantwortung für die Konstitution des sozialen Ereignisses im öffentlichen Raum deutlich: Die Masken- und Kostümträger/innen können das ebenso wenig alleine leisten wie die Spalier stehenden, applaudierenden sowie filmenden und fotografierenden Zuschauer. Nur wenn beide Seiten in angemessener Weise ko-konstruktiv investieren, entsteht für die Dauer des Umzugs die spezifische Qualität eines sozialen Ereignisses im öffentlichen Raum.

Abbildung 2: Fotografieren als ko-konstruktiver Beitrag



4. PRODUKTIVITÄT DES KONZEPTE

Aus multimodaler Perspektive liegt die Produktivität des Konzeptes der Ko-Konstruktion in zwei unterschiedlichen Bereichen:

Zum einen kann aus Sicht der multimodalen Konversationsanalyse die Frage nach den modalitätsspezifischen Möglichkeiten und Restriktionen ko-konstruktiver Praxis beantwortet werden und somit zur Schärfung des Konzeptes beitragen. Man kann beispielsweise systematisch nach multimodalen Ko-Konstruktionen suchen und diese in konzeptreflexiver Weise explizit auf verbale clear cases beziehen. Dabei tauchen eine Reihe interessanter Fragen auf:

- Können Teile einer verbalen Konstruktion auch mit anderen modalitätsspezifischen Ressourcen ko-konstruktiv komplettiert werden? Kann beispielsweise eine gestikulierende Aktivität eine von einem anderen Interaktionsbeteiligten eröffnete verbale Konstruktion vervollständigen?
- Gibt es für visuell wahrnehmbaren Konstruktionsbeteiligungen reservierte Modalitäten?
- Ist eine mischmodale Realisierung nur bei komplettierenden oder auch bei initiiierenden Konstruktionsteilen möglich?
- Welche interaktions- und organisationsstrukturellen Bedingungen müssen für mischmodale Ko-Konstruktionen gegeben sein?
- Ist eine nur visuell wahrnehmbare Komplettierung funktional äquivalent im Vergleich mit einer verbalen Vollform?

- Welche Rolle spielt die Möglichkeiten, verbal gleichsinnig zu komplettieren, in einer andern Modalität (Gestik, Mimik, Blickorganisation) jedoch »Nicht-Gleichsinnigkeit« zu signalisieren?

Dies sind nur einige Fragen, die produktive Irritation entwickeln, wenn man sich von der relativ sicheren Plattform verbaler Ko-Konstruktion entfernt. Ihre Beantwortung würde zu einer Erweiterung des Konzeptes führen, bei denen alle Ausdrucksressourcen gleichwertig Berücksichtigung finden und empirisch hinsichtlich ihrer funktionalen Differenz bzw. Äquivalenz geprüft werden könnten.

Zum anderen sensibilisiert die Beschäftigung mit der ko-konstruktiven Perspektive für die unterschiedlichsten Formen gemeinsam hergestellter sozialer Ordnung, die den oben skizzierten Grundlagen der Sinnproduktion entsprechen: »Gemeinsamkeit der Orientierung«, »Gleichsinnigkeit der Sinnproduktion«, »kollektive Relevanzsetzung thematischer oder pragmatischer Aspekte« sowie eine »wechselseitig akzeptierte koordinative Relevanz«. Ohne eine solche Fokussierung würden viele Formen von Interaktion hinsichtlich ihrer ko-konstruktiven Qualität überhaupt nicht in den Blick geraten.

Diesbezüglich befinde ich mich in vollständiger Übereinstimmung mit Dausendschön-Gay/Gülich/Krafft (i.d.Bd.), wenn sie die egalitäre Relevanz aller Ausdrucksphänomene im Rahmen einer multimodalen Erkenntnisperspektive betonen:

»Im Kern bedeutet eine multimodale Orientierung, dass die Annahme der Situiertheit jedes kommunikativen Ereignisses Beschreibungen fordert, in denen der Beitrag aller verfügbaren Handlungsressourcen zur Sinnkonstitution und ihr Zusammenspiel im interaktiven Geschehen zum Gegenstand der Untersuchungen gemacht werden... « (Dausendschön-Gay/Gülich/Krafft, i.d.Bd.: 35)

LITERATUR

- Collett, Peter/Marsh, Peter (1981): »Patterns of public behavior: Collision avoidance on a pedestrian crossing«, in: *Nonverbal Communication, Interaction and Gesture. Selections from Semiotica (Approaches to Semiotics 41)*, The Hague: Mouton, S. 199-217.
- Hausendorf, Heiko/Schmitt, Reinhold (2013): *Interaktionsarchitektur und Sozialtopografie. Umriss einer raumlinguistischen Programmatik. Arbeitspapiere des UFSP Sprache und Raum (SpuR) Nr. 01*, 59 Seiten, Zürich – Mai 2013 <http://www.spur.uzh.ch/research.html>.
- Lee, John. R. E./Watson, Rodney (1993): »Regards et habitudes des passants. Les arrangements de visibilité de la locomotion«, in: *Annales de la recherche urbaine* 57-58, S. 100-109.
- Ryave, Lincoln A./Schenkein, James N. (1974): »Notes on the art of walking«, in: Roy Turner (Hg.): *Ethnomethodology. Selected readings*, Harmondsworth: Penguin, S. 265-274.
- Schmitt, Reinhold (2006): »Videoaufzeichnungen als Grundlage für Interaktionsanalysen«, in: Sylvia Dickgießer/Ulrich Reitemeier/Wilfried Schütte (Hg.): *"Symbolische Interaktionen"* (Sonderheft Deutsche Sprache 34, 1-2). Berlin, 18-31.

- Schmitt, Reinhold (2007a): »Theoretische und methodische Implikationen der Analyse multimodaler Interaktion«, in: Werner Holly/ Ingwer Paul (Hg.): Medialität und Sprache (Mitteilungen des Deutschen Germanistenverbandes 54/1). Bielefeld: Aisthesis, S. 26-53.
- Schmitt, Reinhold (2007b) (Hg.): Koordination. Analysen zur multimodalen Interaktion (Studien zur Deutschen Sprache Band 38). Tübingen: Narr.
- Schmitt, Reinhold (2012a): »Gehen als situierte Praktik: ‚Gemeinsam gehen‘ und ‚hinter jemandem herlaufen‘«, in: Gesprächsforschung Online-Zeitschrift zur verbalen Interaktion (ISSN 1617-1837), Ausgabe 13, S. 1-44 (www.gesprachsforschung-ozs.de).
- Schmitt, Reinhold (2012b): »Störung und Reparatur eines religiösen Ritus: Die erloschene Osterkerze«, in: Wolfgang Klein/Stephan Habscheid (Hg.): Dinge und Maschinen in der Kommunikation, Sonderheft der Zeitschrift für Literaturwissenschaft und Linguistik 168, S. 62-91.
- Schmitt, Reinhold (in Vorbereitung): »„Frame-Comics“ als Dokumente multimodaler Interaktionsanalysen«, erscheint in: Heiko Hausendorf/Wolfgang Kesselheim/Reinhold Schmitt (Hg.): Interaktionsarchitektur, Sozialtopografie und Interaktionsraum. Tübingen: Narr.
- Schmitt, Reinhold/Dausendschön-Gay, Ulrich (im Druck): Freiraum schaffen im Klassenzimmer: Fallbasierte methodologische Überlegungen zur Raumanalyse. Erscheint in: Arbeitspapiere des UFSP Sprache und Raum (SpuR) Nr. 04, 102 Seiten, Zürich – September 2015, <http://www.spur.uzh.ch/research.html>.